Data Structures, Tree Algorithms and Renewal Theorems

Philippe Robert

INRIA-Rocquencourt

Contents

1. Algorithms & Data Structures

Digital Search Trees (Tries)

Digital Search Trees: Problem

— N elements x_1,. . . , $x_N \in \mathcal{S}$;

— Pb: if $y \in \mathcal{S}$, determine if $y \in \{x_1, \ldots, x_n\}$;

— Minimal nb of operations.

Digital Search Trees: Hash Functions

— $h : \mathcal{S} \longrightarrow [0,1]$;

— For $x \in \mathcal{S}$, $h(x){=}0.X_1^x X_2^x \ldots X_n^x \ldots,$

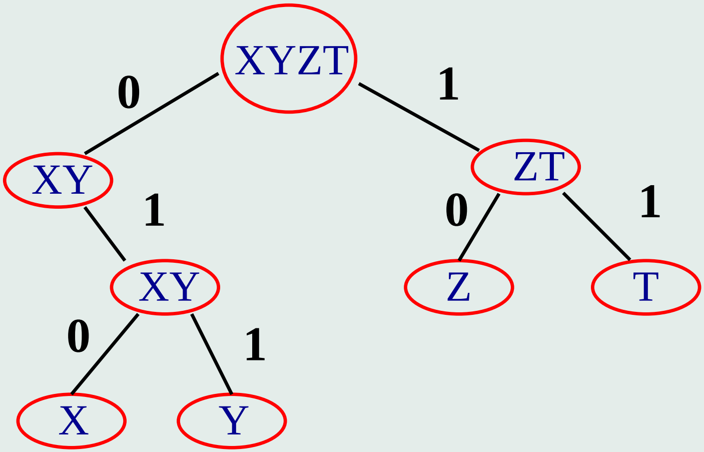
– $X_k^x \in \{0,1\}$, $k \geq 1$,

– **Assumption 1**: $(X_n^x, n{\geq}1)$ independent,

$$\mathbb{P}(X_n^x = 0) = \mathbb{P}(X_n^x = 1) = 1/2.$$

– **Assumption 2**:
$((X_n^x, n{\geq}1), x{\in}\mathcal{S})$ independent,

A digital search tree



h(X)=0.0101.. h(Y)=0.0111..

h(Z)=0.101.. h(T)=0.1100..

Communication Networks

The Context

— Network with N transmitters (stations)
One commom communication channel.

— Two attempts of transmission
on the channel at the same time
$\Rightarrow$ failure.

— A Distributed System:
Centralized Control $\Rightarrow$ Not possible.

Information of a station

A station can probe the channel to detect:

0 —no activity
No attempt on the channel.

1 —success
only one station has transmitted.

2 —failure
$\geq$ 2 stations have tried to transmit.

The channel delivers a ternary information

Transmission Policy

— Based only on the information delivered by the channel

— A station with a message:
Decides to transmit or not.

— All the stations use the same policy.

Simplified Model

— Discrete Time.

— Beginning of Time Unit:
each station decide to transmit or not.

— Duration of a message transmission
= One time unit.

Some history

— Aloha (1968)
Abramson (Hawai)

— Ethernet (1973)
Metcalf and Boggs (Harvard)

— Cambridge Ring (1974)
Cambridge University

— Tree Algorithms (1979)
Capetanakis (MIT)
Tsybakov and Mikhailov (Acad. Sc.Moscow).

The tree algorithm

To each station $S \longrightarrow$ variable "counter" C_S.

Initially $C_S = 0$.

— If $C_S = 0$: try to transmit.

1. Success: end.
2. Collision: flip a coin:
   if head, $C_S = 0$, else, $C_S = 1$.

The tree algorithm (II)

— **If $C_S > 0$, no attempt.**

On the channel:

1. Success or no transmission,
$$C_S \to C_S - 1.$$

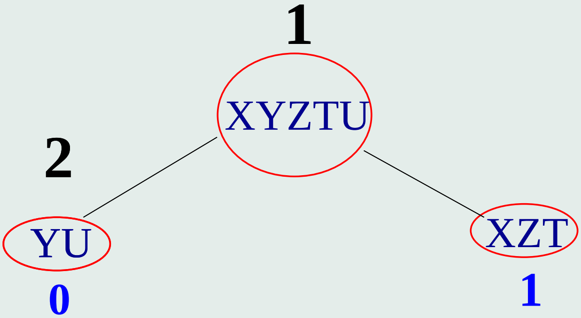
2. Collision $\quad C_S \to C_S + 1.$

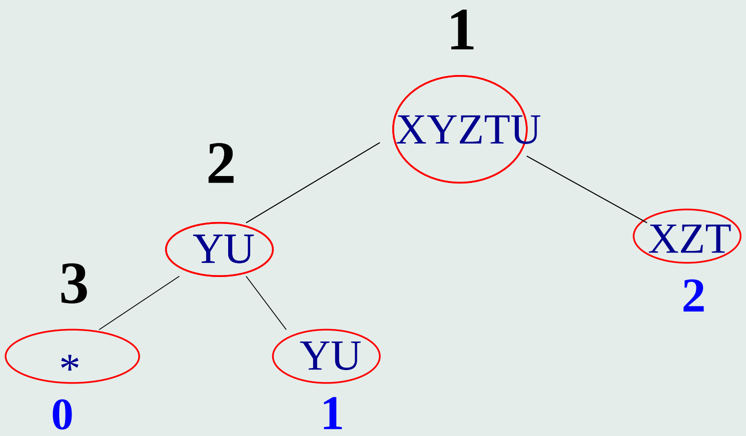
Example

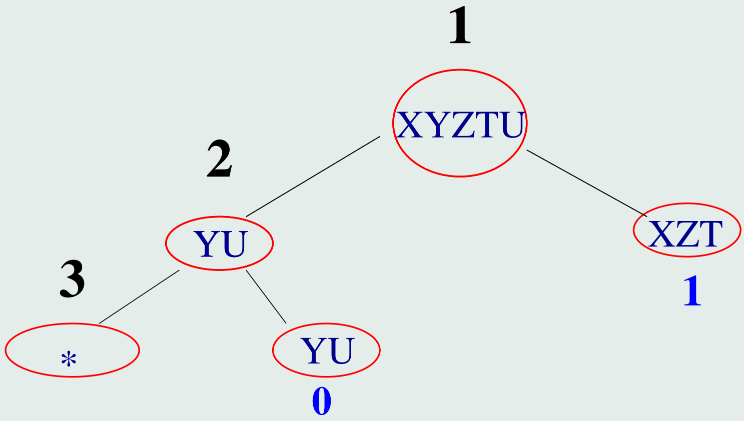
t	1	2	3	4	5	6	7
C	No	No	∅	No	Ok	Ok	No
0	XYZTU	YU		YU	Y	U	XZT
1		XZT	YU	XZT	U	XZT	
2			XZT		XZT		

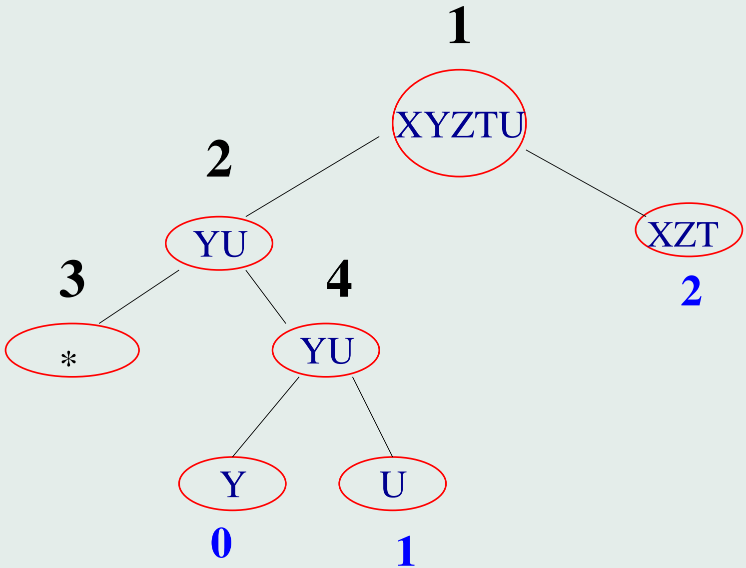
t	8	9	10	11
C	Ok	No	Ok	Ok
0	X	ZT	Z	T
1	ZT		T	

XYZTU

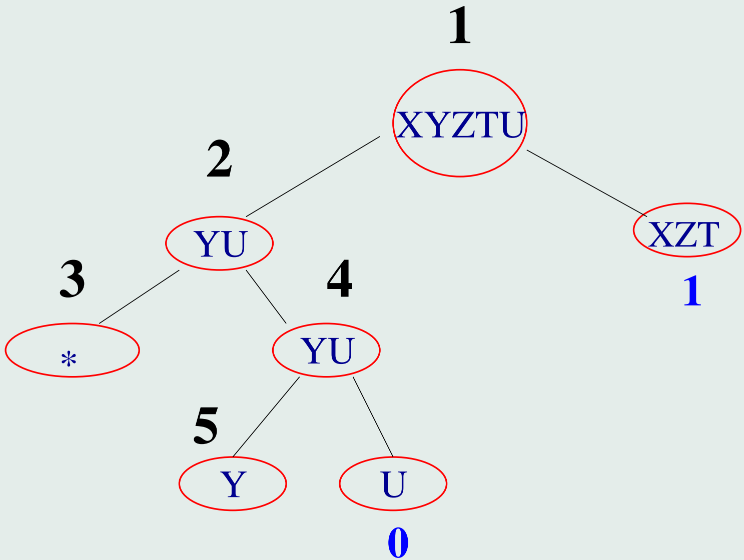

1
XYZTU
2
YU
0
XZT
1

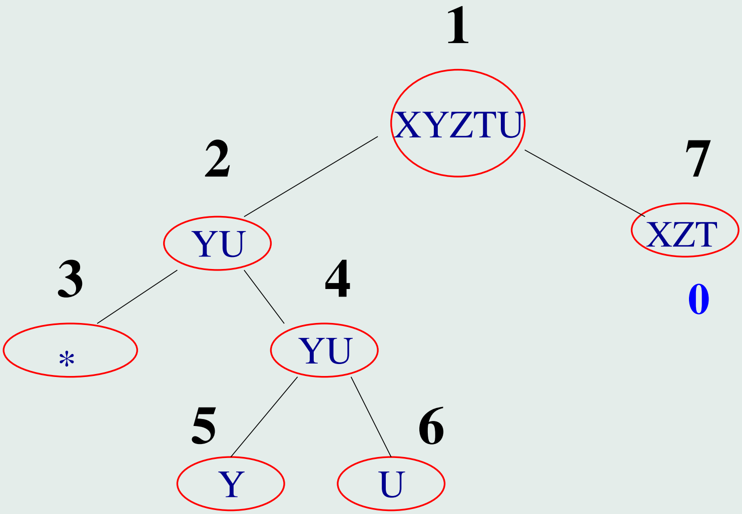

1
XYZTU
2
YU
XZT
2
3
*
YU
0
1

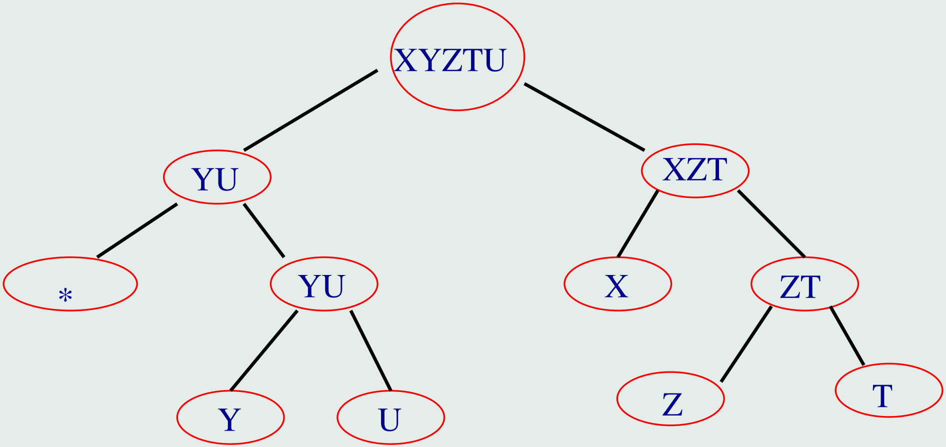

1
XYZTU
2
YU
XZT
1
3
*
YU
0



1
XYZTU
2
YU
XZT
2
3
4
*
YU
Y
U
0
1


1
XYZTU
2
YU
XZT
1
3
4
*
YU
5
Y
U
0


1
XYZTU
2
YU
7
XZT
0
3
*
4
YU
5
Y
6
U



$\mathbf{X = 0.10..,\ Y = 0.010..,\ Z = 0.1110..,}$
$\mathbf{T = 0.1111,\ U = 0.011..}$

Tree Algorithm

A Probabilistic Algorithm
Random numbers are used,
Randomness breaks ties.

Examples of probabilistic Algorithm:

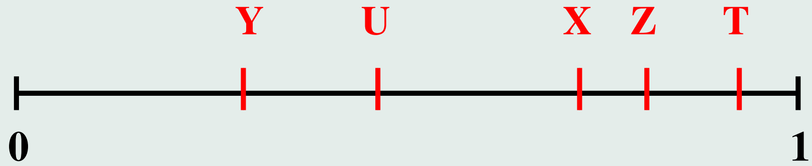
— Routing in Networks.

— Image Analysis.

— Test For Cryptography.

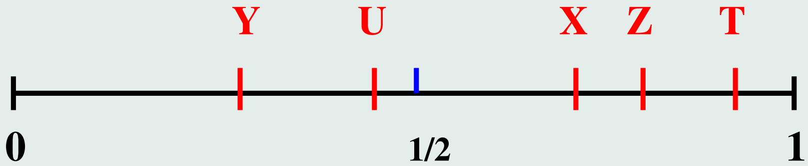
— . . .

An Alternative point of View



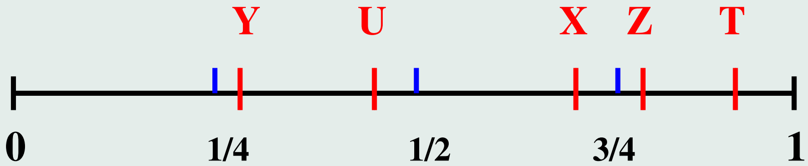
Interval Fragmentation with $\mathbf{5}$ random points in $[\mathbf{0}, \mathbf{1}]$.

Interval Fragmentation



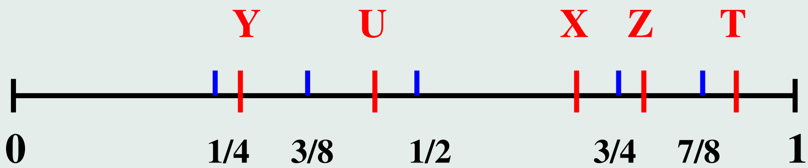
Interval Fragmentation with $\mathbf{5}$ random points in $[0, 1]$.

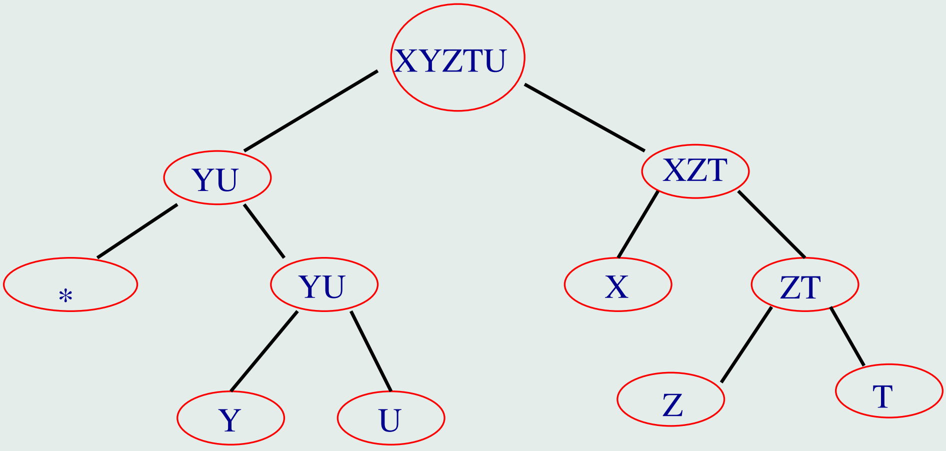
Interval Fragmentation



Interval Fragmentation with $\mathbf{5}$ random points in $[\mathbf{0}, \mathbf{1}]$.

Interval Fragmentation





Statistical Tests

Blood testing

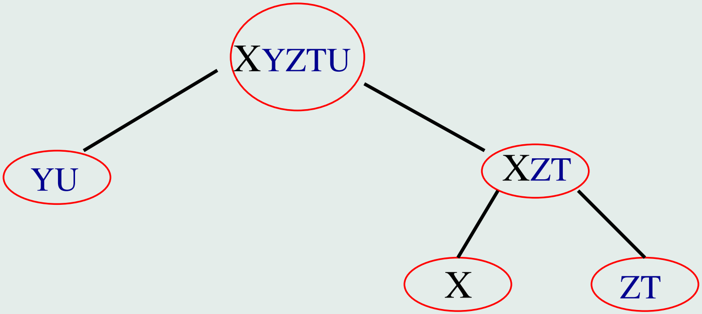
— Very reliable but expensive test.

— Problem: S: set of individuals
how to detect quickly infected ind. in S ?

Algorithm $\mathcal{T}(S)$:
Mix the blood of the elts of S and test.

— If negative: Stop.

— If positive: Split $S = S_1 \cup S_2$
Apply $\mathcal{T}(S_1)$ and $\mathcal{T}(S_2)$.


XYZTU
YU
XZT
X
ZT

Divide and Conquer Algorithms

Divide and Conquer Algorithms

Algorithm $\mathcal{A}(n)$:

— Termination Condition.
If $n < D \longrightarrow$ Stop.

— Tree Structure.
If $n \geq D$, split at random into d groups of size n_1, n_2,... n_d,

$$n_1 + n_2 + \cdots + n_d = n$$

$\Rightarrow$ Apply $\mathcal{A}(n_1)$, $\mathcal{A}(n_2)$, . . . , $\mathcal{A}(n_d)$.

2. Asymptotic Behavior

Asymptotic behavior

How efficient is a tree algorithm ?

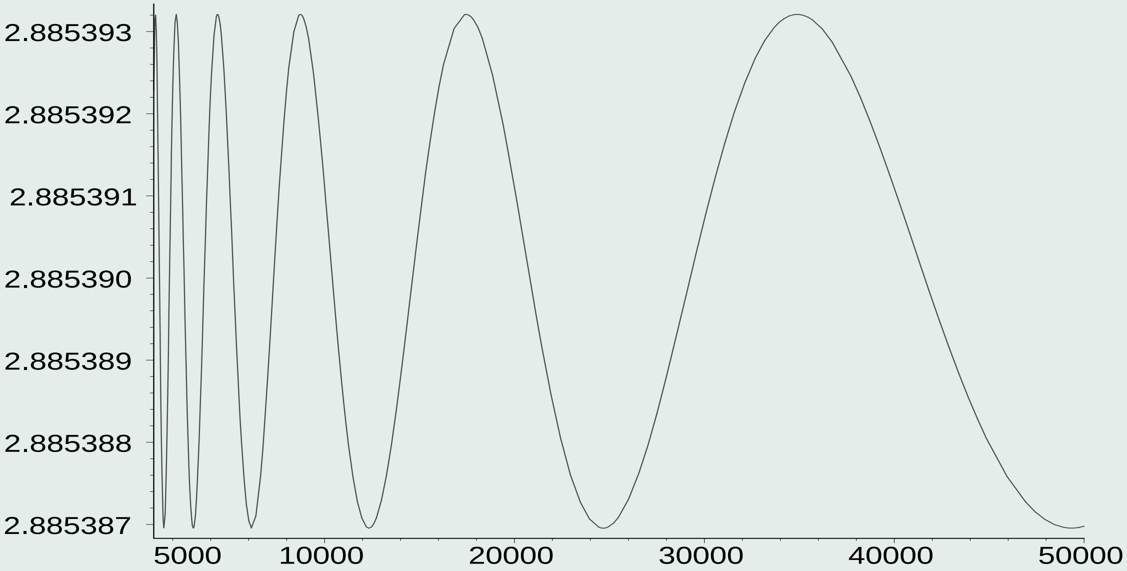
Definition: R_n nb of nodes with n at root.

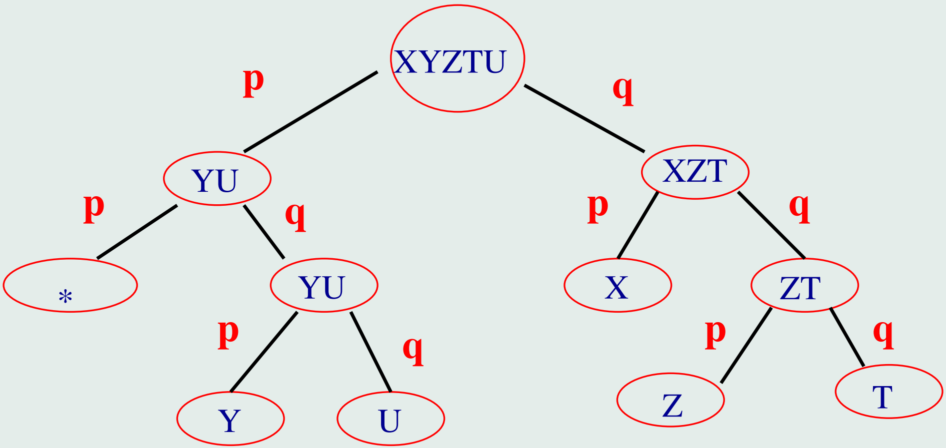
$\dfrac{\mathbb{E}(R_n)}{n}$: Average cost to process 1 item.

Law of Large Numbers: $\displaystyle\lim_{n\to+\infty} \frac{\mathbb{E}(R_n)}{n}$: NO

Pb: the limit does not always exist !

The sequence $n \longrightarrow \mathbb{E}(R_n)/n$ for binary splitting Knuth (1973).





Non-Symmetrical Algorithm

$$p + q = 1$$

Asymptotic behavior

$$\left(\frac{\mathbb{E}(R_n)}{n}\right) : \begin{cases} \text{converges} & \text{if } \frac{\log p}{\log q} \notin \mathbb{Q} \\ \text{oscillates} & \text{otherwise} \end{cases}$$

Literature: Complex analysis methods

Knuth, FLajolet and co-authors.

Analytical Approach

Recurrence Relation

$R_0 = R_1 = 1$. For $n \geq 2$,

$$R_n \stackrel{\text{dist.}}{=} 1 + R_{X_n} + \overline{R}_{n-X_n}$$

with

$$X_n = B_1 + B_2 + \cdots + B_n.$$

(B_i) i.i.d. Bernoulli parameter $1/2$.

$(\overline{R}_n)$ same dist. as (R_n) independent of (R_n)

Poisson Transform

If $r_n = \mathbb{E}(R_n)$

$$r(x) = \sum_{n \geq 0} r_n \frac{x^n}{n!} e^{-x} = \mathbb{E}(r_{N_x}) = \mathbb{E}(R_{N_x})$$

N_x random variable with Poisson dist. with parameter x

$$\mathbb{P}(N_x = n) = \frac{x^n}{n!} e^{-x}$$

Poisson Transform of $(\mathbb{E}(R_n))$

$$r(x) = \sum_{n\geq 0} \mathbb{E}(R_n)\frac{x^n}{n!}e^{-x} = \mathbb{E}(R_{N_x})$$

Since $N_x \sim x$,

$x \longrightarrow \mathbb{E}(R_{N_x})$ and $n \longrightarrow (\mathbb{E}(R_n))$
should have the same behavior at infinity.

But remember that $N_x \sim x + \sqrt{x}\,\mathcal{N}$
$\mathcal{N}$ Gaussian random variable.

A Reminder on Poisson Variables

Thinning of Poisson r.v.:

(B_i) i.i.d. Bernoulli with parameter p

$\mathbb{P}(B_0 = 1) = p$ and $\mathbb{P}(B_0 = 0) = q = 1 - p$

$$\sum_{1}^{N_x} B_i \text{ Poisson with parameter } px.$$

$$N_x - \sum_{1}^{N_x} B_i \text{ Poisson with parameter } qx.$$

Recurrence Relation (II)

$R_0 = R_1 = 1$

If $n \geq 0$ and $Y_n = n - X_n$,

$$R_n \overset{\text{dist.}}{=} 1 + R_{X_n} + \overline{R}_{Y_n} - 2_{\{n \leq 1\}}$$

$$R_{N_x} \overset{\text{dist.}}{=} 1 + R_{X_{N_x}} + \overline{R}_{Y_{N_x}} - 2_{\{N_x \leq 1\}}$$

Thinning of Poisson r.v.: if $r(x) = \mathbb{E}(R_{N_x})$

$$r(x) = r(px) + r(qx) + 1 - 2(1 + x)e^{-x}$$

Analytic Approach: Mellin Transform

r fn on $\mathbb{R}_+$, Mellin Transform of r:

$$r^*(s) = \int_0^{+\infty} r(x) x^{s-1}\, dx$$

generally defined in a strip $\mathcal{S}$ of $\mathbb{C}$.

Some Properties:

— $g : x \to r(\mu x) \Rightarrow g^*(s) = \mu^{-s} r^*(s)$

— Asymptotic behavior of r at $+\infty$ determined by poles of r^* on right of $\mathcal{S}$

Flajolet and co-authors.

Analytic Approach: A Summary

1. Poisson Transform: $(r_n) \longrightarrow r(x)$

$$r(x) = r(px) + r(qx) + 1 - 2(1+x)e^{-x}$$

2. Mellin:

$$r^*(s) = (p^{-s} + q^{-s})r^*(s) + h^*(s)$$

3. Inversion of Mellin: $r(x)$ as $x \longrightarrow +\infty$.

4. Inversion of Poisson: r_n as $n \longrightarrow +\infty$.

3. A Probabilistic approach

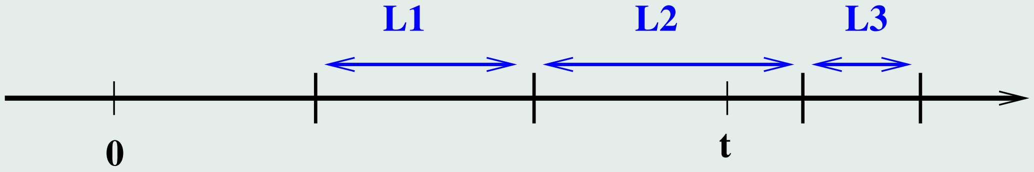
Renewal theorems

Some History

Blackwell (1948):

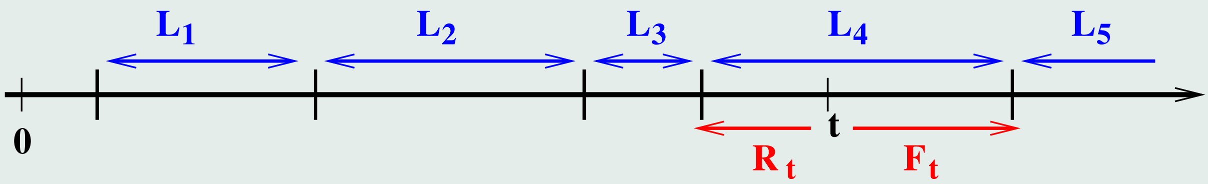
— A light bulb last two years in average.

— How many are necessary for ten years ?



Breiman, Feller, Lindvall, . . .

General Framework



(L_i) i.i.d. non-negative random variables.

— $U(a, b)$: average number of points in $[a, b]$,
U: Renewal measure,

$$\text{for } h > 0 \qquad \lim_{t \to +\infty} U(t, t+h)?$$

— Behavior of (R_t, F_t) as $t \to +\infty$?

Renewal Theorem

Non-Lattice Case: $\forall\delta > 0$, $\mathbb{P}(L \in \delta\mathbb{N}) < 1$

$$\lim_{t\to+\infty} U(t, t+h) = \frac{h}{\mathbb{E}(L_1)}$$

$$(R_t, F_t) \stackrel{\text{dist.}}{\longrightarrow} (R_\infty, F_\infty) :$$

$$\mathbb{E}(f(R_\infty, F_\infty)) = \frac{1}{\mathbb{E}(L_1)} \mathbb{E}\left(\int_0^{L_1} f(u, L_1 - u)\, du\right)$$

F_t has density $x \to \mathbb{P}(L_1 \geq x)/\mathbb{E}(L_1)$

Proofs

— **Renewal Equation:**

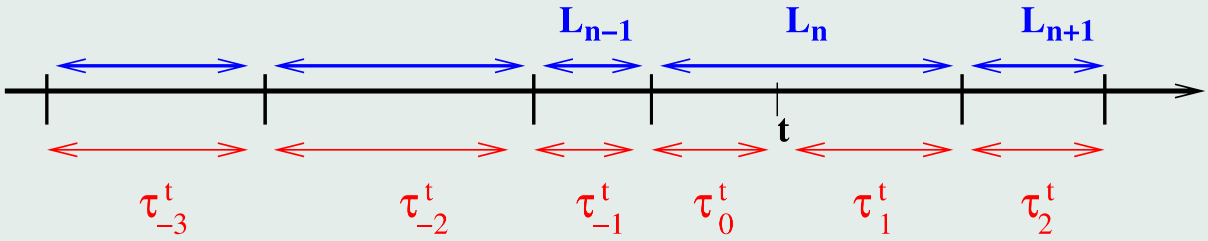
$$U(0,t) = 1 + \int_0^t U(0, t-u) L(du)$$

$\Rightarrow$ Fourier Analysis (Feller).

— **Coupling:**

Lindvall, Athreya and Ney, . . .

A Point Process Point of View



Renewal Theorem:

$$(\tau_i^t, i \in \mathbb{Z}) \xrightarrow{\text{dist.}} (L_i^*, i \in \mathbb{Z})$$

Stationary renewal process.

1. $(L_i^*, i \leq -1)$, $(L_i^*, i > 1)$, i.i.d. dist. as L_1;

2. $(L_0^*, L_1^*) \overset{\text{dist.}}{=} (R_\infty, F_\infty)$.

Lattice Case

If $\mathbb{P}(L_1 \in \delta\mathbb{N}) = 1$ and $\mathbb{P}(L_1 = \delta) > 0$:

— $(\tau_i^t, i \in \mathbb{Z})$ does not converge as $t \to +\infty$

— for $h > 0$,

$$(\tau_i^{h+n\delta}, i \in \mathbb{Z})) \overset{n\to+\infty}{\longrightarrow} (L_i^{*h}, i \in \mathbb{Z})$$

— Periodic Behavior

$$(L_i^{*h}, i \in \mathbb{Z}) \overset{\text{dist.}}{=} (L_i^{*(h+\delta)}, i \in \mathbb{Z})$$

A Reminder on Poisson Process

If $0 \leq t_1 \leq \cdots \leq t_n \leq \cdots$ **with** $(t_{n+1} - t_n)$ **i.i.d.**

$$\mathbb{P}(t_{n+1} - t_n \geq x) = \exp(-x),$$

If $N_x = \mathrm{card}\{k : t_k \leq x\}$,

N_x **Poisson r.v. with parameter** x

$$\mathbb{P}(N_x = p) = \frac{x^p}{p!} e^{-x}$$

Back to the recurrence Relation

$$R_n \overset{\text{dist.}}{=} 1 + R_{X_n} + \overline{R}_{n-X_n} - 2 \times 1_{\{n \leq 1\}}$$

with

$$X_n = B_1 + B_2 + \cdots + B_n.$$

$(B_i) \in \{0, 1\}$ i.i.d. Bernoulli parameter p.

$(\overline{R}_n)$ same dist. as (R_n) independent of (R_n)

Poisson Transform

$$R_n \overset{\text{dist.}}{=} 1 + R_{X_n} + \overline{R}_{n-X_n} - 2 \times 1_{\{n \le 1\}}$$

$$\mathbb{E}(R_{N_x}) = \mathbb{E}(R_{N_{px}}) + \mathbb{E}(R_{N_{qx}}) + 1 - 2\mathbb{P}(t_2 \ge x)$$

t_2 second point of Poisson process (N_x)

If $r(x) = \mathbb{E}(R_{N_x}) - 1$

$$r(x) = r(px) + r(qx) + 2\mathbb{P}(t_2 < x)$$

Poisson Transform: a probabilistic rewriting

$$r(x) = r(px) + r(qx) + 2\mathbb{P}(t_2 < x)$$

If $A_1 \in \{p, q\}$ r.v. such that $\mathbb{P}(A_1 = p) = p$

$$r(x) = \mathbb{E}\left(\frac{r(A_1 x)}{A_1}\right) + 2\mathbb{E}\left(1_{\{t_2 < x\}}\right)$$

Poisson Transform: a probabilistic rewriting

$$r(x) = r(px) + r(qx) + 2\mathbb{P}(t_2 < x)$$

If $A_1 \in \{p, q\}$ r.v. such that $\mathbb{P}(A_1 = p) = p$

$$r(x) = \mathbb{E}\left(\frac{r(A_1 x)}{A_1}\right) + 2\mathbb{E}\left(1_{\{t_2<x\}}\right)$$

$$= \mathbb{E}\left(\frac{r(A_2 A_1 x)}{A_2 A_1}\right) + 2\mathbb{E}\left(\frac{1}{A_1} 1_{\{t_2<xA_1\}}\right) + 2\mathbb{E}\left(1_{\{t_2<x\}}\right)$$

Poisson Transform: a probabilistic rewriting

$$r(x) = r(px) + r(qx) + 2\mathbb{P}(t_2 < x)$$

If $A_1 \in \{p, q\}$ r.v. such that $\mathbb{P}(A_1 = p) = p$

$$r(x) = \mathbb{E}\left(\frac{r(A_1 x)}{A_1}\right) + 2\mathbb{E}\left(\mathbf{1}_{\{t_2<x\}}\right)$$

$$= \mathbb{E}\left(\frac{r(A_2 A_1 x)}{A_2 A_1}\right) + 2\mathbb{E}\left(\frac{1}{A_1}\mathbf{1}_{\{t_2<xA_1\}}\right) + 2\mathbb{E}\left(\mathbf{1}_{\{t_2<x\}}\right)$$

$$= 2\sum_{k=0}^{+\infty}\mathbb{E}\left(\frac{1}{\prod_1^k A_i}\mathbf{1}_{\{t_2<x\prod_1^k A_i\}}\right)$$

Poisson Transform

$$\mathbb{E}\left(R_{N_x}\right) = 1 + 2\sum_{k=0}^{+\infty} \mathbb{E}\left(\frac{1}{\prod_1^k A_i}\mathbf{1}_{\{t_2 < x\prod_1^k A_i\}}\right)$$

$$= 1 + 2\sum_{k=0}^{+\infty} \mathbb{E}\left(\sum_{n=2}^{+\infty}\frac{1}{\prod_1^k A_i}\mathbf{1}_{\{t_2 < x\prod_1^k A_i, N_x = n\}}\right)$$

Given $\{N_x = n\}$ the n points of Poisson proc.: n i.i.d. uniformly dist. r.v. on $[0, x]$

$U_{2,n}$: second smallest value of n i.i.d. uniform r.v. on $[0, 1]$

$$\mathbb{P}\left(t_2 \leq \frac{x}{\mathcal{A}}, N_x = n\right)$$
$$= \mathbb{P}\left(t_2 \leq x/\mathcal{A} \middle| N_x = n\right) \frac{x^n}{n!} e^{-x}$$

$$\mathbb{P}\left(t_2 \leq x/\mathcal{A} \middle| N_x = n\right) = \mathbb{P}\left(U_{2,n} \leq 1/\mathcal{A}\right)$$

Probabilistic de-Poissonization

$$\mathbb{E}\left(R_{N_x}\right) = \sum_{n\geq 0} \mathbb{E}(R_n)\frac{x^n}{n!}\,e^{-x}$$

$$= 1 + 2\sum_{k=0}^{+\infty}\sum_{n=2}^{+\infty}\mathbb{E}\left(\frac{1}{\prod_1^k A_i}\mathbf{1}_{\{t_2<x\prod_1^k A_i, N_x=n\}}\right)$$

$$\mathbb{E}\left(R_{N_x}\right) = \sum_{n\geq 0} \frac{x^n}{n!}\,e^{-x}$$

$$+2\sum_{n=2}^{+\infty}\sum_{k=0}^{+\infty}\mathbb{E}\left(\frac{1}{\prod_1^k A_i}\mathbf{1}_{\{U_{2,n}<\prod_1^k A_i\}}\right)\frac{x^n}{n!}\,e^{-x}$$

An Associated Random Walk

$$\mathbb{E}(R_n) = 1 + 2\mathbb{E}\left(\sum_{k\geq 0} \frac{1}{\prod_1^k A_i} 1_{\{U_{2,n} < \prod_1^k A_i\}}\right), \; n \geq 2.$$

$$\frac{\mathbb{E}(R_n) - 1}{2n} = \mathbb{E}\left(\sum_k e^{-\log n + \sum_{i=1}^k -\log(A_i)} \times 1_{\left\{-\sum_1^k \log(A_i) \leq -\log U_{2,n}\right\}}\right)$$

The Use of Renewal Theorem

$$L_i = -\log A_i \qquad U_{2,n} \sim 1/n$$

$$\Delta_n = \mathbb{E}\left(\sum_k e^{-\left(\log n - \sum_{i=1}^k L_i\right)} \mathbf{1}_{\left\{\sum_1^k L_i \le \log n\right\}}\right)$$

$$\Delta_n \sim \mathbb{E}\left(\sum_{k\le 0} e^{-\sum_{i=k}^0 \tau_i^{\log n}}\right) \sim \mathbb{E}\left(\sum_{k\le 0} e^{-\sum_{i=k}^0 L_i^*}\right)$$

if L_0 **non-lattice** $\Leftrightarrow \log p/\log q \notin \mathbb{Q}$

Renewal Theorems

— Dist. of $-\log A$ non-lattice: $\log p/\log q \notin \mathbb{Q}$.
Continuous Renewal Theorem.
Convergence of $\mathbb{E}(R_n)/n$.

— Dist. of $-\log A$ lattice: $\log p/\log q \in \mathbb{Q}$.
Discrete Renewal Theorem.
Periodic Fluctuations of $\mathbb{E}(R_n)/n$.

General Case

Algorithm $\mathcal{A}(n)$

— $n < D \Rightarrow$ STOP.

Otherwise:

— Take a r.v. $G \in \mathbb{N}$; Branching variable

— Take a random probability vector $\mathcal{V} = (V_1, \ldots, V_G)$, $V_1 + \cdots + V_G = 1$;

Weights on arcs

— Split n into G subgroups $(n_1, \ldots n_G)$ randomly, according to vector $\mathcal{V}$.

— Apply $\mathcal{A}(n_1)$, . . . , $\mathcal{A}(n_G)$.

Splitting Measure

Probability Measure on $[0, 1]$:

$$\int_0^1 f(x)\,\mathcal{W}(dx) = \mathbb{E}\left(\sum_{i=1}^{G} V_{i,G} f(V_{i,G})\right).$$

Theorem. [Mohamed and R. (2005)] If

$$\int_0^1 \frac{|\log(y)|}{y} \mathcal{W}(dy) < +\infty,$$

— if $\log \mathcal{W}$ non-lattice,

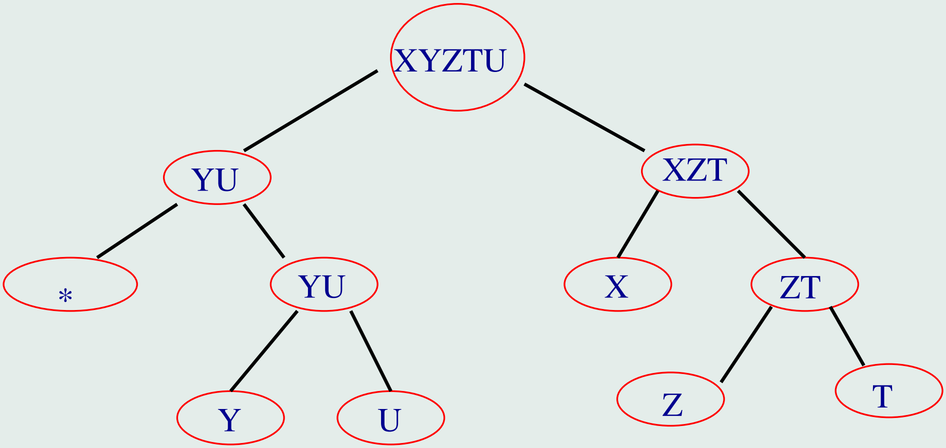
$$\lim_{n \to +\infty} \mathbb{E}(R_n)/n = \frac{\mathbb{E}(G)}{(D-1) \int_0^1 |\log(y)| \mathcal{W}(dy)}.$$

— $\log \mathcal{W}$ lattice: $\mathbb{E}(R_n)/n \sim F(\log n/\lambda)$

$$F(x) = C \int_0^{+\infty} \exp\left(-\lambda\left\{x - \frac{\log y}{\lambda}\right\}\right) \frac{y^{D-2}}{(D-1)!} e^{-y}\, dy$$

4. Stationary Digital Trees

Tree Structure



$$\mathbf{X = 0.10.., \ Y = 0.010.., \ Z = 0.1110..,}$$

$$\mathbf{T = 0.1111, \ U = 0.011..}$$

Sequences

— Words of a dictionary, book;

— DNA sequences;

An element: $X = (X_k, k \in \mathbb{Z}) \in \{0,1\}^{\mathbb{Z}}$
X_k: stationary random variables;

Stationarity:

$$(X_k, k \in \mathbb{Z}) \stackrel{\text{Dist.}}{=} (X_{k+p}, k \in \mathbb{Z})$$

Stationary sequences

— $X = (X_k, k \in \mathbb{Z}) \in \{0,1\}^{\mathbb{Z}}$
a stationary sequence;

— Each item x has a sequence (X_k^x)
with same distribution as X;

— (X_k^x), $x \in S$ are independent;

X_i^x "coin" (possibly) used by x at level i;

— R_n size of the tree with n items.

Asymptotic behavior of $\mathbb{E}(R_n)/n$?

Functional Analysis Approach

Context

— A function $\phi : [0,1] \to [0,1]$;

— A Partition $[0,1] = \cup_i I_i$;

— $X_n = p$ if $\phi^{(n)}(X_0) \in I_p$.

Sequence (X_n) generated by iteration of a function.

Baladi, Bourdon, Clément, Flajolet, Vallée, . . .

Functional Analysis Approach

Methods

— Ruelle's Transfer Operator:

$$\mathcal{G}_s(f)(z) = \sum_{i=1}^{r} \phi_i(z)^s f \circ \phi_i(z)$$

— Spectral analysis:
Study the dominant eigenvalue $\lambda(s)$ of the operator $\mathcal{G}_s$ for $s \sim 1$.

Back to General case: Cost Function

$$C_n(f) = \sum_{\alpha \in \Sigma^*} f(np_\alpha)$$

Σ^* finite vectors: $\bigcup_{n=0}^{+\infty} \quad \{0,1\}^n$

p_α proba. of vector $\alpha = (\alpha_1, \ldots, \alpha_n) \in \Sigma^*$

$p_\alpha = \mathbb{P}(X_1 = \alpha_1, \ldots, \ldots, X_n = \alpha_n \mid X_k, k < 0),$

Exemple:

$$R_n = \sum_{\alpha \in \Sigma^*} (1 - (1 + np_\alpha)(1 - p_\alpha)^n) \sim \sum_{\alpha \in \Sigma^*} \int_0^{np_\alpha} u e^{-u} du$$

Probabilistic rewriting of cost function

$$C_n(f) = \sum_{\alpha\in\Sigma^*} f(np_\alpha)$$

$$= \sum_{k\geq 1} \sum_{\substack{\alpha\in\Sigma^* \\ |\alpha|=k}} \frac{f(np_\alpha)}{p_\alpha} p_\alpha$$

$$= \sum_{k\geq 1} \mathbb{E}\left(\frac{f(np_{\kappa_k})}{p_{\kappa_k}}\right)$$

κ_k projection on k first coord.

Rewriting of cost function: Fubini's Theorem

$$C_n(f) = \sum_{k\geq 1} \mathbb{E}\left(\frac{f(np_{\kappa_k})}{p_{\kappa_k}}\right)$$
$$= \mathbb{E}\left(\sum_{k\geq 1}\frac{1}{p_{\kappa_k}}\int_0^{+\infty} 1_{\{u\leq np_{\kappa_k}\}} f'(u)\,du\right)$$
$$= \int_0^{+\infty} \mathbb{E}\left(G_n(u)\right) f'(u)\,du,$$

with

$$G_n(u) = \sum_{k\geq 1}\frac{1}{p_{\kappa_k}} 1_{\{u\leq np_{\kappa_k}\}}.$$

A rough estimation

Shannon's Theorem: $-\log p_{\kappa_k} \sim kH$, a.s.
H entropy.

$$
\begin{aligned}
G_n(u) &= \sum_{k\geq 1} \frac{1}{p_{\kappa_k}} 1_{\{u \leq n p_{\kappa_k}\}} \\
&= \sum_{k\geq 1} e^{-\log p_{\kappa_k}} 1_{\{-\log p_{\kappa_k} \leq \log(n/u)\}} \\
&\text{"}\sim\text{"} \sum_{k\geq 1} e^{kH} 1_{\{k \leq \log(n/u)/H\}} \\
&= \frac{e^{\lceil \log(n/u)/H \rceil H} - 1}{e^H - 1} \sim \frac{n}{u(e^H - 1)}.
\end{aligned}
$$

G_n of order n.

More Rigor

$$p_{\kappa_n(\omega)} = \mathbb{P}(X_n{=}\omega_n, \ldots, X_0 = \omega_0 \mid X_k{=}w_k, k < 0)$$

$$= \prod_{i=1}^{n} \mathbb{P}(X_i{=}\omega_i \mid X_i{=}\omega_i, \ldots, X_0{=}\omega_0, X_k{=}\omega_k, k < 0)$$

$$-\log p_{\kappa_n} = \sum_{i=1}^{n} h \circ \theta^i(\omega)$$

θ: shift on sequences;

$$h(\omega) = -\log \mathbb{P}(X_0 = \omega_0 \mid X_k = \omega_k, k < 0).$$

h entropy function, $H = \mathbb{E}(h)$

More Rigor (II)

$$-\log p_{\kappa_n} = \sum_{i=1}^{n} h \circ \theta^i(\omega)$$

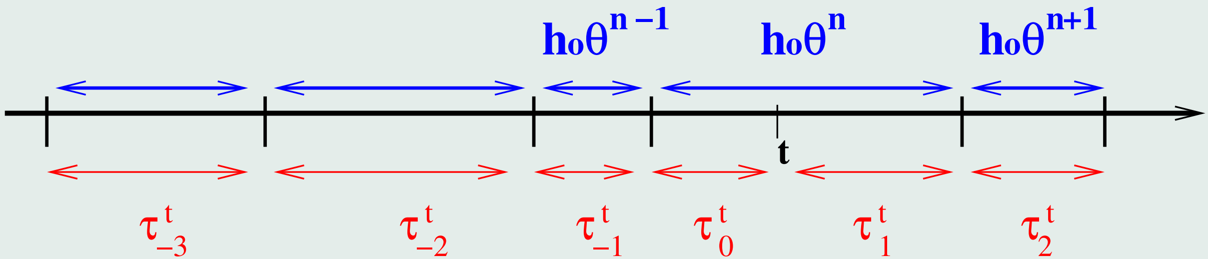
$$G_n(u) = \sum_{k \geq 1} \frac{1}{p_{\kappa_k}} 1_{\{u \leq n p_{\kappa_k}\}}$$

$$\frac{G_n(u)}{n} = \frac{1}{u} \sum_{k \geq 1} \exp\left[-\left(\log(n/u) - \sum_{1}^{k} h \circ \theta^i\right)\right]$$

$$1_{\left\{\sum_{1}^{k} h \circ \theta^i \leq \log(n/u)\right\}}$$

$\Rightarrow$ Renewal Theorem Context.

Renewal Theory for Stationary Sequences



Renewal Theorem: $(h \circ \theta^n)$ stationary ergodic,

$$(\tau_i^t, i \in \mathbb{Z}) \xrightarrow{\text{dist.}} ?$$

Renewal Theory for Stationary Sequences (II)

Guivarc'h and Hardy (1988), Lalley (1989)

— **Application:** Limit theorems for counting measures of periodic orbits of flows.

Blanchard (1976) Delasnerie and Neveu (1977)
Equivalence between

— Renewal thm. for sequence $(h \circ \theta^n)$;

— Mixing property of special flow under h.

— The variable h is non lattice.

A convergence result

Entropy function

$$h(x) = -\log(P(X_0 = x \mid X_{-1}, X_{-2}, \ldots)).$$

Theorem [R. (2005)] If $H = \mathbb{E}(h(X_0))$, and the distribution of $h(X_0)$ is not lattice,

$$\lim_{n \to +\infty} \frac{C_n(f)}{n} = \frac{1}{H} \int_0^{+\infty} \frac{f'(u)}{u} \, du.$$

Application

$$\lim_{n \to +\infty} \frac{\mathbb{E}(R_n)}{n} = \frac{1}{H}.$$

The end